



Bachelor Level / First Year/ First Semester/ Science
Computer Science and Information Technology Stat. 103
(Probability and Statistics)

Full Marks: 60
Pass Marks: 24
Time: 3 hours.

Candidates are required to give their answers in their own words as far as practicable.
All notations have the usual meanings.

Group A

Attempt any Two:

(2×10=20)

- ✓ 1. Write the algebraic computation expressions for mean and standard deviation based on a given sample $x_1, x_2, \dots, x_n$. Why they are important in statistics? Write down their properties compute the mean and standard deviation from the following scores of 10 students.

45	55	35	60	55	55	65	40	45	35
----	----	----	----	----	----	----	----	----	----

2. Explain the terms – sample space and an event of a random experiment. State the classical and the statistical definition of probability. Which of the two definitions is most useful in statistics and why? If A, B and C are events of a sample space such that $P(A) = 0.5$, $P(A \cap C) = 0.2$, $P(A \cap B^c \cap C^c) = 0.1$, and $P(A \cap B \cap C) = 0.05$, using Venn-diagram find $P(A \cap B)$.

Venn-diagram find $P(A \cap B)$.

- ✓ 3. A large company wants to measure the effectiveness of newspaper advertising media on sale promotion of its products. A sample of 22 cities with approximately equal populations is selected for study. The sales of the product (Y) in thousand Rs and the level of newspaper advertising expenditure (X) in thousand Rs are recorded for each of the 22 cities (n) and the recorded sum, sum of square, and sum of cross product of X and Y are summarized below.

$$\Sigma Y = 26953, \quad \Sigma X = 660, \quad \Sigma Y^2 = 35528893, \quad \Sigma X^2 = 22700, \quad \& \quad \Sigma YX = 851410$$

Using the above summary results:

- (a) Compute correlation coefficient r between X and Y, and coefficient of determination.
9 (b) Fit a simple linear regression model of Y on X using least square method and interpret the estimated slope regression coefficient.

Group B

Answer any eight questions:

(8×5=40)

- ✓ 4. Write down the role of probability theory in statistics with suitable examples.

5. Explain discrete and continuous random variables with suitable examples. Suppose a continuous random variable X has the density function

$$f(x) = \begin{cases} kx^2 & \text{if } -1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) value of the constant k , and (b) find $E(X)$.

6. Suppose that X and Y have joint density function

$$f(x, y) = \begin{cases} 4xy & \text{if } 0 < x < 1 \text{ and } 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find (a) marginal densities of X and Y , (b) $P(X \leq 0.3)$ and (c) $P(Y \geq 0.5)$.

7. In a binomial distribution with parameters n and p derive the mean and variance of the distribution.

8. If $X_1, X_2, \dots, X_n$ are n independent Poisson random variables with common mean λ , derive the maximum likelihood estimator of λ . Show that the estimator is unbiased for λ .

9. If $Z_1, Z_2, \dots, Z_n$ are n independently distributed standard normal variates, what is the distribution of $\sum_{i=1}^n Z_i^2$? Write down the density function of $\sum_{i=1}^n Z_i^2$ and mention its properties.

10. If a continuous random variable X has exponential distribution with density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

For $h > 0$, prove that $P(X > t + h | X > t) = P(X > h)$, and hence prove that $P(X > t + h) = P(X > t) \times P(X > h)$.

11. If a random variable X is normally distributed with a mean of 120 and a standard deviation of 12. Compute the following probabilities: (a) $P(X > 130)$, (b) $P(X < 115)$, and (c) $P(110 < X < 130)$.

12. A manufacturer of TV sets claims that the average life of its picture tubes is at least 10 years. A sample survey of 100 of the picture tubes showed an average of 9.6 years and a standard deviation of 2.6 years. Do these results cast doubt on the claim of manufacturer? Answer this question by setting appropriate null and alternative hypotheses and testing the null hypothesis at 5% level of significance.

13. The Chamber of Commerce claims that the mean carbon dioxide level of air pollution is no more than 4.9 ppm. A random sample of 16 readings resulted mean equal to 5.6 ppm and standard deviation equal to 2.1 ppm. Assuming that the carbon dioxide level is normally distributed, is there evidence against the Chamber of Commerce's claim? Answer the query by setting appropriate null and alternative hypotheses and testing the null hypothesis at 5% level of significance.